# Regulation of $H_2O_2$ Stress-responsive Genes through a Novel Transcription Factor in the Protozoan Pathogen *Entamoeba histolytica*[*][S]

**Richard J. Pearson**[‡], **Laura Morf**[‡], and **Upinder Singh**[‡§1]

*From the ‡Division of Infectious Diseases, Department of Internal Medicine, and the §Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305*

**Background:** How gene expression is regulated in response to oxidative stress is unknown in *Entamoeba histolytica*.
**Results:** Motif AAACCTCAATGAAGA, enriched in promoters of $H_2O_2$-responsive genes, specifically binds amoebic protein EHI_108720.
**Conclusion:** EHI_108720 is a transcription factor that mediates up-regulation of gene expression in response to elevated $H_2O_2$ levels.
**Significance:** Determining the molecular basis of $H_2O_2$ stress response is critical to understanding parasite virulence.

Outcome of infection depends upon complex interactions between the invading pathogen and the host. As part of the host's innate immune response, the release of reactive oxygen and nitrogen species by phagocytes represents a major obstacle to the establishment of infection. The ability of the human parasite *Entamoeba histolytica* to survive reactive oxygen and nitrogen species is central to its pathogenic potential and contributes to disease outcome. In order to define the transcriptional network associated with oxidative stress, we utilized the MEME and MAST programs to analyze the promoter regions of 57 amoebic genes that had increased expression specifically in response to $H_2O_2$ exposure. We functionally characterized an $H_2O_2$-regulatory motif (HRM) ([1]AAACCTCAATGAAGA[15]), which was enriched in these promoters and specifically bound amoebic nuclear protein(s). Assays with promoter-luciferase fusions established the importance of key residues and that the HRM motif directly impacted the ability of $H_2O_2$-responsive promoters to drive gene expression. DNA affinity chromatography and mass spectrometry identified EHI_108720 as an HRM DNA-binding protein. Overexpression and down-regulation of EHI_108720 demonstrated the specificity of EHI_108720 protein binding to the HRM, and overexpression increased basal expression from an $H_2O_2$-responsive wild-type promoter but not from its mutant counterpart. Thus, EHI_108720, or HRM-binding protein, represents a new stress-responsive transcription factor in *E. histolytica* that controls a transcriptional regulatory network associated with oxidative stress. Overexpression of EHI_108720 increased parasite virulence. Insight into how *E. histolytica* responds to oxidative stress increases our understanding of how this important human pathogen establishes invasive disease.

Hydrogen peroxide elicits a robust transcriptional response in multiple organisms (1–7) and, along with other reactive oxygen species and nitrogen species (ROS[2] and RNS), targets multiple cellular components (8). In response, microbes have developed a wide range of defense mechanisms to either directly tackle ROS and RNS or repair the damage that they cause (9–11). Understanding how this response is coordinated at a transcriptional level provides important information regarding the ability of microorganisms to survive in hostile environments. In *Escherichia coli* and yeast, the transcription factors OxyR and YAP1, respectively, have been identified as the principal players in coordinating the transcriptional response to hydrogen peroxide (12, 13). These transcription factors are directly impacted by elevated hydrogen peroxide levels and display altered DNA binding specificity (OxyR) (14, 15) or elevated protein levels in the nucleus (YAP-1) (16), resulting in up-regulation of multiple stress response genes (reviewed in Ref. 17).

*Entamoeba histolytica*, a protozoan parasite, is an important human pathogen that must survive changing oxygen tensions and ROS in order to establish infection. Ninety percent of individuals infected with *E. histolytica* remain asymptomatic, whereas 10% develop a potentially lethal, invasive disease (18). The basis of this variable disease presentation is not fully understood but is probably due in part to the virulence potential of different parasite strains. Both virulent and non-virulent strains of *E. histolytica* have been identified (19), and comparative analyses of the proteome and transcriptome have identified multiple virulence determinants (20–22). One particularly striking difference is the increased expression in the virulent strain of the surface molecule peroxiredoxin, which degrades hydrogen peroxide (23). It has been demonstrated that virulent

[2] The abbreviations used are: ROS, reactive oxygen species; HRM, hydrogen peroxide-responsive motif; HRM-BP, HRM-binding protein; RNS, reactive nitrogen species.

*E. histolytica* strains survive exposure to oxidative stress better than avirulent strains, in part due to the presence of peroxiredoxin (21).

On the transcriptome level, microarray studies demonstrated that exposure to sublethal quantities of $H_2O_2$ or dipropylenetriamine-NONOate (a nitric oxide releaser) results in greater changes in transcript levels in a virulent than in a nonvirulent amoebic strain (2). The percentage of genes regulated by these compounds in the pathogenic strains is larger and the magnitude of changes observed in individual genes is higher than that observed in the non-pathogenic strains (2). The majority of the known factors, including peroxiredoxin, that protect against ROS and RNS are more highly expressed in the virulent strains, but because they are already expressed at robust levels, they do not significantly alter their expression levels in response to stress. This suggests that the virulent strains of *E. histolytica* utilize transcriptional networks in response to ROS or RNS to regulate the expression of either novel protective factors or factors required in other aspects of increased virulence.

Transcriptional regulation remains a poorly understood aspect of *E. histolytica* biology, and only a few transcription factors and their corresponding DNA binding motifs have been characterized (reviewed in Ref. 24). Of those transcription factors that have been well characterized, most were originally selected due to sequence similarity to known factors, such as EhMyb10 or the EhTBP (25, 26). However, some unique transcription factors have been successfully identified in *E. histolytica*; the best characterized of these is URE3-BP, an EF-hand domain-containing protein that is regulated by changes in calcium levels (27, 28).

In this study, we identify a new *E. histolytica* transcription factor that plays a role in coordinately regulating gene expression in response to hydrogen peroxide exposure. We employed a bioinformatics approach to identify an $H_2O_2$-responsive motif (HRM) that was enriched within promoters of genes upregulated following exposure to stress. Our functional studies demonstrated that this motif specifically binds to an amoebic nuclear protein(s), and mutation of this motif resulted in altered gene expression. We utilized a combined DNA affinity chromatography and mass spectrometry approach to identify the HRM-binding protein (HRM-BP), EHI_108720. This protein specifically interacts with the HRM, and manipulation of HRM-BP expression levels altered basal expression and stress responsiveness of an $H_2O_2$-responsive promoter. These data represent the first steps in elucidating the transcriptional network responsible for coordinating changes in gene expression following $H_2O_2$ exposure in the important human pathogen *E. histolytica*.

## EXPERIMENTAL PROCEDURES

*Microarray Data*—Microarray data using custom-generated arrays from Affymetrix identified 184 genes that are up-regulated in the *E. histolytica* strain HM-1:IMSS in response to a 1-h exposure to 1 mM $H_2O_2$ (2). For the purposes of this study, this set of genes was reduced by the removal of genes that are upregulated in response to NO (2) or heat shock (29). Reannotation of the genome (30, 31) resulted in a further reduction and

resulted in 57 promoters being used for analysis in this study. The gene promoters used for this study are listed in supplemental Table 1.

*Databases*—The *E. histolytica* genome was downloaded from Pathema along with the location of all predicted open reading frames (ORFs) (download date, January 8, 2009). This enabled the retrieval of −300 to −1 nucleotide regions relative to the predicted translation start site for each ORF. For each predicted promoter region, sequence was retrieved irrespective of the location of surrounding predicted ORFs.

*DNA Motif Identification*—The MEME and MAST programs were used as described previously (29, 32). In brief, the MEME and MAST programs were obtained from the University of California San Diego (33, 34). MEME was performed with the command line: *-dna -mod zoops -minw 6 -maxw 10 -minsites 5 -nmotfs 20* or *-minw 10 -maxw 14*. This set of commands identifies 20 motifs that must have zero or one occurrence in each promoter and occur a minimum of five times within the total number of promoters. The motifs found would have a length between 6 and 10 nucleotides or between 10 and 14 nucleotides. The MAST program was utilized to determine the total number of occurrences of each motif in the promoter sequence databases. The command line arguments used for this purpose were as follows: *-ev 500 -remcorr*. This allows sequences with an *e* value of less than 500 and examines each pair of motifs and removes motifs that have a correlation coefficient of greater than 0.7. The hypergeometric distribution was used to determine the significance of enrichment for each motif identified. Motifs with a *p* value of less than 0.01 were determined to be significantly enriched within the promoters of the 57 $H_2O_2$-up-regulated genes (supplemental Table 2). Sequence logos were generated using WebLogo (35).

*Isolation of Cytoplasmic and Nuclear Enriched Fractions*—*E. histolytica* HM-1:IMSS were grown axenically in TYI-S-33 medium at 36.5 °C as described previously (36). Nuclear extraction was performed using previously published methods (29, 37) with the following modifications. Amoeba were harvested and resuspended in Buffer A (10 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl and 0.6% IGEPAL) with protease inhibitors 1 $\mu$M leupeptin, 1 $\mu$M E-64-d, and 1× HALT protease inhibitor mixture (Pierce) and incubated on ice for 20 min. Samples were briefly vortexed and centrifuged for 10 min, at 1000 × *g* at 4 °C. The supernatant was removed and stored at −80 °C and represents the cytoplasmic fraction. The nuclear pellet was then washed in 500 $\mu$l of Buffer A (without IGEPAL), followed by centrifugation for 10 min at 150 × *g* and 4 °C. The supernatant was removed and the nuclei were resuspended in 150 $\mu$l of buffer C (20 mM HEPES, pH 7.9, 420 mM NaCl, 1 mM EDTA, and 1 mM EGTA) supplemented with the same protease inhibitor mixture. Samples were incubated on ice for 30 min prior to centrifugation (20 min, 4 °C, 18,000 × *g*). The supernatant was removed, and aliquots were frozen on dry ice prior to being stored at −80 °C. Protein distribution was assessed by Western blot analysis. Cytoplasmic and nuclear fractions (30 $\mu$g of total protein/sample) were separated by 12.5% SDS-PAGE, and blots were probed with antibodies to histone H3 (ab1791, Abcam), actin (691001, MP Biomedicals), and Myc tag (9B11, Cell Signaling).

*EMSA Analysis*—Complementary overlapping probes were designed with the motif of interest centrally located, annealed, and labeled using [$\alpha$-$^{32}$P]dATP and Klenow fragment (Invitrogen) as described previously (29). Binding reactions took place in a total volume of 20 $\mu$l and included EMSA binding buffer (10 mM Tris-HCl, pH 7.9, 50 mM NaCl, 1 mM EDTA, 3% glycerol, 0.05% milk powder, and 0.05 mg of bromphenol blue), 5 $\mu$g of nuclear extract, 2 $\mu$g of poly(dI-dC), and 50 fmol of labeled probe. Supershift assays were performed by the addition of 1 $\mu$g of either a specific $\alpha$-Myc antibody (sc-40X, Santa Cruz Biotechnology, Inc.) or control $\alpha$-$\beta$-actin antibody (sc-47778X, Santa Cruz Biotechnology, Inc.), and competition assays were performed with cold oligonucleotide probes provided in 100-fold excess. The supershift and competition assays required the addition of antibody or competitor to the binding reaction prior to the addition of the labeled probe (30-min room temperature incubation for antibody and 10 min for competitor). Upon the addition of the labeled probe, the final reaction mix was incubated for 30 min at room temperature, and samples were loaded onto a non-denaturing polyacrylamide gel and run for 3 h. Upon completion, the gel was fixed, dried, and exposed to a phosphor screen. Gels were imaged using software from Molecular Dynamics. The oligonucleotides used in EMSA (mutations underlined) were as follows: consensus probe, 5′-acgtcagtAAACCTCAAT-GAAGAcagt (sense) and 5′-acgtactgTCTTCATTGAGGTTT-actg (antisense); mutant 1, 5′-acgtcagtAAACCTCA<u>GATAAAG</u>-cagt (sense) and 5′-acgtactg<u>CTTTATC</u>TGAGGTTTactg (antisense); mutant 2, 5′-acgtcagt<u>CAATCCA</u>AATGAAGAcagt (sense) and 5′-acgtactgTCTTCATT<u>TGGATTG</u>actg (antisense); mutant 3, 5′-acgtcagtAAACCT<u>AAC</u>TGAAGAcagt (sense) and 5′-acgtactgTCTTCA<u>GTT</u>AGGTTTactg (antisense); mutant 4, 5′-acgtcagtAA<u>A</u>CCTCAATGAAGAcagt (sense) and 5′-acgtactg-TCTTCATTGAG<u>TG</u>TTactg (antisense); mutant 5, 5′-acgtcagt-AAACCTCAA<u>GT</u>AAGAcagt (sense) and 5′-acgtactgTCTT<u>AC</u>-TTGAGGTTTactg (antisense).

The EMSA probes used for promoter analysis were as follows: EHI_029340, 5′-acgtcagtAAACCTCAATGTTGAcagt (sense) and 5′-acgtactgTCAACATTGAGGTTTactg (antisense); EHI_134960, 5′-acgtcagtAAACCTCAATAAACTcagt (sense) and 5′-acgtactgAGTTTATTGAGGTTTactg (antisense); EHI_176810, 5′-acgtcagtAAACCTCATTGTTCAcagt (sense) and 5′-acgtactgTGAACAATGAGGTTTactg (antisense).

*Reporter Gene Analysis*—Promoter regions of genes EHI_029340, EHI_134960, and EHI_176810 were amplified from *E. histolytica* HM-1:IMSS genomic DNA using the following primers: EHI_029340, 5′-ACA<u>AAGCTT</u>TCATGTTGAGAAT-TAAGTGGTTCA-3′ (forward) and 5′- ACA<u>CCTAGG</u>TGGTG-TTGTGTTTGACATGAAT-3′ (reverse); EHI_134960, 5′-ACA-<u>AAGCTT</u>TTCTTAATTAGATTCATGGTTATTC-3′ (forward) and 5′- ACA<u>CCTAGG</u>TTTATCACTTTTTTTTCATAATGAA-3′ (reverse); EHI_176810, 5′- ACA<u>AAGCTT</u>TTTGAGTTATATATT-CAAGAAAGAA-3′ (forward) and 5′- ACA<u>CCTAGG</u>AAGTG-CATATCTTATCATTTTATG-3′ (reverse). Approximately 500 nucleotides were selected for each promoter region. All reverse primers incorporated the first 15 nucleotides downstream of the start codon for the given gene, which has previously been demonstrated to be important for optimal gene expression in *E. histolytica* (38). Promoters were cloned into the pKT-Luc vector using

HindIII and AvrII restriction sites underlined in the primers. Site-directed mutagenesis was performed to insert the desired mutations into the DNA motif using the QuikChange site-directed mutagenesis protocol (Stratagene). All vectors were sequenced prior to being utilized in functional analyses.

Transient transfection of *E. histolytica* HM-1:IMSS parasites was achieved as follows. Amoebae were plated into a 35-mm tissue culture dish 24 h prior to transfection. Medium was removed, and cells were washed with 4 ml of M199 and then covered with 1.8 ml of M199 plus 15% fetal bovine serum. Transfection was achieved using 10 $\mu$g of experimental plasmid DNA and 1 $\mu$g of a control *Renilla* luciferase plasmid, incubated at room temperature for 10 min with 30 $\mu$l of SuperFect (Qiagen) in M199 medium to a total volume of 200 $\mu$l. This mixture was added dropwise to the amoeba, and dishes were incubated at 36.5 °C for 3 h. Subsequently, the dishes were iced for 10 min to release the parasites, and the amoeba were transferred to a standard 16-ml culture tube and resuspended in *E. histolytica* culture medium. Following a 16 h incubation, transfected amoebae were chilled on ice for 5 min, and the cells were pelleted at $1000 \times g$ for 5 min. Luciferase assays were accomplished using the Dual-Luciferase reporter assay system (Promega) following the manufacturer's protocol. Pelleted cells were resuspended in 100 $\mu$l of cell lysis buffer (plus 1 $\mu$M leupeptin, 1 $\mu$M E-64-d, and 1× HALT protease inhibitor mixture (Pierce)) and incubated on ice for 10 min. Samples were vortexed briefly prior to centrifugation (30 s, maximum speed), and supernatants were transferred to fresh tubes. Of each sample, 10 $\mu$l was assessed for expression of both firefly and *Renilla* luciferase, and firefly luciferase results were normalized using the *Renilla* luciferase readings. Stress was applied by the addition of 1 mM $H_2O_2$ to parasites for 1 h prior to isolation of whole cell extracts, which were then processed as described above.

*Enrichment and Identification of EHI_108720*—Protein EHI_108720 was isolated using a DNA affinity chromatography technique described previously (39, 40). The following double-stranded wild type and mutant biotinylated probes were designed: WT P1, 5′-biotinylated ACGTCAGTACACAA-CAAACCTCAATGAAGACAGTACGT-3′; mutant P1, 5′-biotinylated ACGTCAGTACACAACAGACCATAACTA-AAGCAGTACGT-3′. Biotinylated probes were fixed to streptavidin-Sepharose beads (Cell Signaling) and incubated with 300 $\mu$g of crude nuclear extract (precleared with non-biotinylated mutant probe and poly(dI-dC)) for 30 min at room temperature in a 1.5-ml tube. Beads were washed twice with 100 mM KCl buffer (20 mM HEPES-KOH (pH 7.9), 100 mM KCl, 1 mM DTT, 1 mM EDTA, 0.01% Nonidet P-40, and 15% glycerol). The first 100 mM KCl wash was kept and is the equivalent of the flow-through sample obtained when performing the DNA affinity chromatography protocol exclusively on columns. Beads were resuspended with 1 ml of 100 mM KCl buffer and loaded onto a Mobicol minicolumn (Boca Scientific). The columns were washed with 200 mM KCl buffer (6 ml), and bound proteins were eluted with 500 mM KCl buffer (0.8 ml). All samples were concentrated using an YM-10 Microcon column (Millipore) (to ~30 $\mu$l), and 5 $\mu$l was used for EMSA analysis, with the remainder of the eluted samples from the WT and mutant columns being sent for mass spectrometry.

*Direct Isolation of the Protein(s) of Interest*—Three EMSA binding reactions were performed as described above. Two reactions contained $^{32}$P-labeled HRM consensus oligonucleotide probe, and the third contained unlabeled probe. Samples were run on a polyacrylamide gel with the cold sample being flanked by the two radiolabeled samples. The gel was exposed to film overnight and then aligned on top of the film, allowing the location of the three samples to be determined. A small section of gel was excised that corresponded to the location of the cold probe-protein complex and was analyzed by mass spectrometry (MS).

*Mass Spectrometry*—Samples were submitted to the MS facility at Stanford University and processed in the following manner.

*Gel Slice Sample*—In-gel digestion was done using Promega MS grade trypsin overnight as reported previously (41), with the addition of the acid-labile surfactant ProteaseMAX (Promega). Prior to digestion, the gel slices were cut into ~1 × 1-mm cubes, reduced with 5 mM DTT, and alkylated with acrylamide. Peptides were extracted and dried prior to reconstitution and analysis.

*DNA Affinity Chromatography Samples*—The protein eluates were digested using an acetone precipitation step, reconstituted in 8 M urea/ProteaseMAX, 50 mM ammonium bicarbonate, reduced, and alkylated followed by overnight digestion using trypsin at a 1:100 protease/protein ratio.

Nano-reversed phase HPLC was done using an Eksigent 2D NanoLC system (Eksigent, Dublin, CA) with buffer A consisting of 0.1% formic acid in water and buffer B consisting of 0.1% formic acid in acetonitrile. A fused silica column self-packed with Duragel C18 (Peeke, Redwood City, CA) matrix was used with a linear gradient from 2% B to 40% B at a flow rate of 600 nl/min. The nano-HPLC was interfaced with a Bruker/Michrom Advance Captive spray source for nano-electrospray ionization into the mass spectrometer. The mass spectrometer was an LTQ Orbitrap Velos (Thermo Fisher Scientific), which was set in data-dependent acquisition mode to perform MS/MS on the top 12 most intense multiply charged cations. The .RAW data were searched using Sequest on a Sorcerer platform against the Uniprot database. Data were validated and visualized using Scaffold software.

*Genetic Manipulation of EHI_108720*—For gene overexpression, the entire coding region for EHI_108720 was PCR-amplified from genomic DNA using the primers 5'-ACAC<u>CCGGG</u>ATGGAAGAAGATCACGAT-3' (forward) and 5'-ACA<u>CTCGAG</u>TTAATGATAAAATGTTCCTTTACC-3' (reverse). Myc-tagged EHI_108720 was generated by cloning the full open reading frame into the *E. histolytica* plasmid pKT3M, downstream of the cysteine synthase promoter and 3× Myc tag (cloned into the SmaI and XhoI restriction sites). The resultant construct was transfected into *E. histolytica* HM-1:IMSS using the SuperFect (Qiagen) protocol described above. To establish stable transfection, 20 μg of plasmid DNA was used, and 24 h after transfection, drug selection was started with 1 μg/ml G418 added and increased to 3 μg/ml 48 h later. Medium was exchanged as needed until the tube was confluent with stably transfected amoeba. At this stage, the amoebae were passaged,

and G418 was gradually increased to a final concentration of 12 μg/ml.

For knockdown of the gene, we used a new RNAi-based approach that we have recently developed.[3] Briefly, the entire EHI_108720 coding sequence was cloned downstream of the EHI_197520 small RNA "trigger" fragment in the pKT-3M backbone, using the SmaI and XhoI sites. Stable cell lines were established using the same methodologies as described above. Knockdown efficiency was assessed using RT-PCR (described below).

*RNA Isolation and RT-PCR*—RNA was isolated from untransfected HM-1:IMSS and EHI_108720 overexpression and knockdown cell lines using the mirVana kit (Ambion), and RNA concentration was measured using a Nanodrop spectrophotometer. RT-PCR was performed as described previously (42). Primers used in RT-PCR were as follows: EHI_108720, 5'-GTGG-TATGGGTCAAGCTGGT-3' (forward) and 5'-CAGTGTCTT-GCCAAGTTCCA-3' (reverse); EHI_199600, 5'-ACGTCATGC-TGAATTTGCTG-3' (forward) and 5'-CCTTTAAGCCCAGC-CTTTCT-3' (reverse); EHI_176810, 5'-CCGGTTTGTGATCC-AAGTTT-3' (forward) and 5'-TTGGAATTTGATTAATTCTT-TGAGT-3' (reverse). cDNA samples were subjected to 29 cycles of PCR amplification.

*Cell Monolayer Destruction Assays*—Assays were performed as described previously (43). Briefly, $5 \times 10^4$ trophozoites were placed on a confluent CHO cell monolayer, centrifuged for 5 min at $50 \times g$, and incubated for 75 min at 37 °C. Cells were fixed for 10 min with 4% ultrapure formaldehyde, washed twice with PBS, stained with 0.1% methylene blue (OmniPur), diluted in 10 mM borate buffer (pH 8.7), and washed three times with the same buffer. The dye was extracted by adding 1 ml of 0.1 M HCl at 37 °C for 30 min. In order to measure the extracted dye, samples were diluted 1:10 with PBS, and the absorbance was read at 650 nm in a spectrophotometer.

## RESULTS

### Identification of Conserved Motifs in the Promoter Regions of $H_2O_2$-responsive Genes

In a previous study, Vicente *et al.* (2) identified 184 genes that were up-regulated ≥2-fold in response to hydrogen peroxide in *E. histolytica*. In order to identify promoter motifs in $H_2O_2$-responsive genes, we narrowed this list of genes to those that were specifically up-regulated by the $H_2O_2$ stress. Thus, genes that were also up-regulated by nitric oxide (NO) and/or heat shock were removed from this list (82 genes for NO, 13 genes for heat shock, and 21 genes for both NO and heat shock); in total, 116 genes were removed (2, 29). Additionally, because the microarray data were initially published, the *E. histolytica* genome has been reannotated (31), resulting in a further small decrease in the number of promoters to be analyzed. Ultimately, 57 promoters of genes that were up-regulated specifically by $H_2O_2$ (supplemental Table 1) were examined using the MEME program (33) to identify conserved motifs that could be important in the coordinated regulation of gene expression in response to hydrogen peroxide stress. MEME analysis of pro-
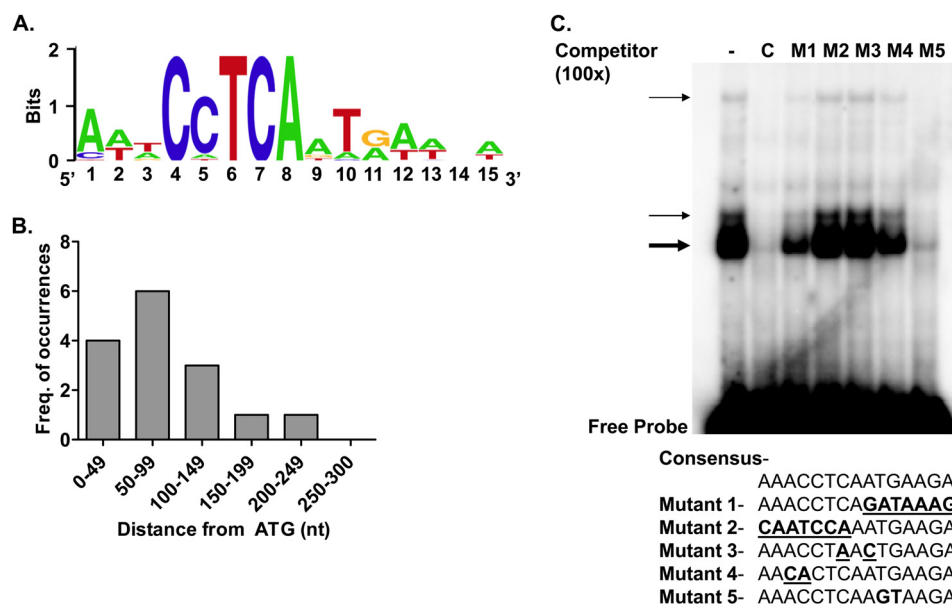
---

FIGURE 1. **Identification of a conserved motif in promoters of *E. histolytica* genes up-regulated by hydrogen peroxide stress.** *A*, MEME and MAST analysis identified the AAACCTCAATGAAGA motif as being enriched in this subset of $H_2O_2$-responsive promoters. *B*, the motif is predominantly located within 60–140 nucleotides of the start codon. *C*, EMSA analysis demonstrates that amoebic nuclear protein(s) bind to this motif. Competition assays, using a 100-fold excess of cold competitor, demonstrated the specificity of this interaction and identified C7 and A9 as residues essential for protein binding. The *arrows* indicate major bands that exhibit specific binding. The sequence logo was generated using WebLogo (35).

moter regions was performed using 300 nucleotides upstream of the start codon and searching for motifs of either 6–10 or 10–14 nucleotides in length. Subsequently, we utilized the MAST program (34) to identify all occurrences of the identified motifs in the promoters of the *E. histolytica* genome to determine which motifs were significantly enriched in the $H_2O_2$-responsive subset (supplemental Table 2).

### The HRM Specifically Binds Amoebic Nuclear Protein(s)

Motifs identified were prioritized based on (i) enrichment of the motif within the subset of promoters relative to the number of occurrences in the entire promoter set (as determined using the hypergeometric distribution ($p < 0.01$)), (ii) highly conserved sequence within the promoters of co-regulated genes, and (iii) relatively conserved position in relation to the start codon. The motif CCTCAAT fulfilled all of the criteria listed above. It was identified 15 times within the 57 promoters studied and was enriched within 100 nucleotides of the start codon in 75% of the promoters in which it was present (Fig. 1*B*). A consensus sequence was obtained by identifying a 15-nucleotide region in each of the promoters that incorporated the CCTCAAT motif (supplemental Table 3). This enabled the identification of nucleotides enriched within the core motif flanking regions. A consensus motif of AAACCTCAAT-GAAGA was established for this HRM (Fig. 1*A*) and subsequently tested for its ability to bind nuclear protein(s). EMSA analysis demonstrated that the consensus probe binds to protein(s) present in crude nuclear extract (Fig. 1*C*) and that this interaction decreased in the presence of a 100-fold excess of cold consensus probe, indicating specificity of the interaction.
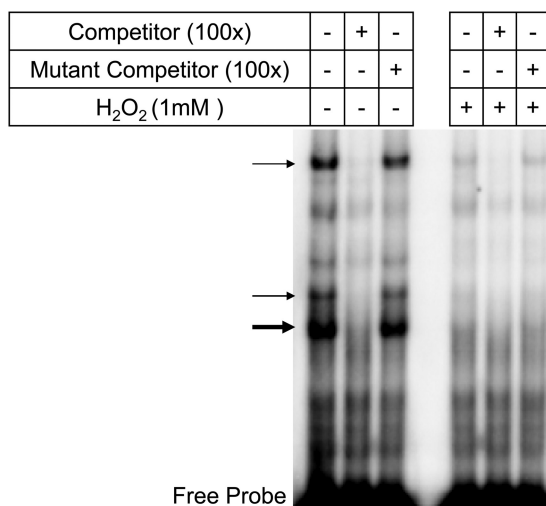
To determine which residues of the consensus probe were critical to binding, several mutant (*M*) probes were used as cold competitors (Fig. 1*C*). The 5′-region of the motif is more relevant to DNA-protein binding because probe M2 could not

compete against the wild-type labeled probe, whereas probe M1 could partially compete. Smaller changes, consisting of the switching of nucleotides at positions 7 and 9 (probe M3) and positions 3 and 4 (probe M4), demonstrated that the maintenance of the ACCTCAAT region was critical for DNA-protein binding. In particular, the reversal of positions C7 and A9 (probe M3) was sufficient to prevent this mutant probe from competing against the labeled consensus probe. Similar changes at the 3′-end (nucleotides 10 and 11; probe M5) had no effect on the ability to compete with the labeled probe when compared with the cold consensus competitor.

To determine if the binding of protein(s) to the HRM was altered significantly under stress conditions, EMSA was performed using nuclear extract isolated from amoebae that had been treated with 1 mM $H_2O_2$ for 1 h (the same stress as used in the transcriptome analyses) (2) (Fig. 2). Nuclear extract from parasites exposed to 1 mM $H_2O_2$ had decreased binding to the labeled HRM probe. However, binding to a control probe was unchanged, demonstrating that loss of binding was specific to the HRM-interacting protein(s) and not due to the quality of the nuclear extract under stress conditions (data not shown). Thus, the data suggest that under stress conditions, either the abundance of the DNA-binding protein(s) decreases, the protein or protein complex is no longer present in the nucleus, or the protein(s) has been modified in such a way that prevents binding to the HRM.

### Functional Characterization of the HRM

*Binding to Promoter-specific Motifs from Three $H_2O_2$-responsive Genes*—In order to determine if the HRM was functionally relevant, three genes were selected that were up-regulated in response to $H_2O_2$ and whose promoters contain the HRM (Table 1). Genes EHI_029340 and EHI_176810 have low basal expression levels with 3- and 8-fold increase, respectively, fol-

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Competitor (100x) | - | + | - | | - | + | - |
| Mutant Competitor (100x) | - | - | + | | - | - | + |
| H₂O₂ (1mM ) | - | - | - | | + | + | + |

Free Probe

**FIGURE 2. Binding to the HRM motif decreases in nuclear extracts from hydrogen peroxide-treated parasites.** Amoebic cultures were either left untreated or exposed to 1 mM H₂O₂ (1 h) prior to isolation of nuclear fractions. EMSA analysis was performed using 5 μg of crude nuclear extract per binding reaction. Specificity of binding was demonstrated using a 100-fold excess of cold wild type or mutant competitor. Competitor probe was AAACCTCAATGAAGA, and mutant competitor probe was AAACCTAACTGAAGA. The *arrows* indicate major bands that exhibit specific binding.

lowing H₂O₂ stress. Gene EHI_134960 is highly expressed at base line and increases 2-fold in response to stress. Despite the differences in basal and regulated mRNA levels and small variations in nucleotide composition, the HRM motif in each of the three promoters bound nuclear protein(s) in a manner similar to the consensus motif (Fig. 3). The binding to labeled probe was competed by excess consensus probe and by promoter-specific probe. Binding to the labeled probe was not affected by the presence of excess mutant competitor, where the nucleotide composition and order only differed by the switching of nucleotides 7 and 9. For all three promoter motifs, H₂O₂ treatment resulted in reduced binding of nuclear protein(s), similar to what was seen with the consensus probe.

*Mutation of the HRM in H₂O₂-responsive Gene Promoters Affects Reporter Gene Expression Levels*—To ascertain the functional impact of the motif, each promoter was cloned upstream of a luciferase reporter gene. Additionally, the motifs in each promoter were mutated using site-directed mutagenesis to switch the nucleotides at positions 7 and 9, which prevented DNA-protein interaction in EMSA. All three wild type (WT) promoters were able to drive robust luciferase expression, with promoter EHI_134960 driving luciferase expression between 3- and 6-fold higher than the other two promoters (Fig. 4A); of note, this matched the greater mRNA abundance for this gene as noted by microarray analysis (Table 1). Interestingly, although mutations in all promoters affected reporter gene expression, the outcome was variable. In two genes (EHI_029340 and EHI_176810), mutation of HRM decreased luciferase expression; in EHI_134960, mutation of HRM increased luciferase expression.

The different results seen between promoters could be attributed to the fact that DNA-protein binding and downstream changes in transcriptional response are context-dependent. The EMSA results are only representative of binding to a short oligonucleotide probe, and the presence of the entire promoter region probably influences gene expression, as monitored in luciferase assays. Taken together, these data highlight the complexity of transcriptional regulation and that the protein or protein complex that binds to the HRM may act differently depending upon additional factors. There is precedence for transcription factors in *E. histolytica* having a dual role, with transcription factor URE3-BP positively influencing expression from the Gal/GalNAc-inhibitable lectin *hgl5* promoter but negatively affecting gene expression from the ferredoxin 1 promoter (37, 44).

*The Promoter of EHI_176810 Responds to H₂O₂ Stress*—We next sought to determine if the WT and mutant promoters were responsive to H₂O₂ stress. Eighteen hours following transient transfection, *E. histolytica* trophozoites were exposed to 1 mM H₂O₂ for 1 h, and luciferase levels were compared with controls without H₂O₂ exposure. Of the three promoters, only the WT promoter for EHI_176810 demonstrated increased luciferase levels in response to stress (Fig. 4B). The EHI_176810 promoter with the mutated HRM did not drive increased luciferase expression after stress, indicating that it was the HRM specifically that was responsible for directing increased gene expression in response to stress. Luciferase expression driven by promoters from EHI_029340 and EHI_134960 did not change in response to H₂O₂. The reason for the difference in response to oxidative stress for the three promoters is unclear. It is of note, however, that the highest level of mRNA change seen on the microarray data is for gene EHI_176810 (Table 1). The change at the luciferase protein level is less than the 8-fold increase seen at the mRNA level for EHI_176810; thus, it is possible that the smaller changes seen at the mRNA level for genes EHI_029340 and EHI_134960 cannot be recapitulated in the luciferase assay, which relies upon protein activity.

### Identification of the HRM-interacting Protein

*Enrichment of DNA-binding Proteins by DNA Affinity Chromatography*—Having identified the HRM as being enriched in H₂O₂-responsive genes and demonstrated that it can alter luciferase expression levels from three promoters, we utilized the HRM in DNA affinity columns to identify the protein(s) that interacts with it. Biotinylated oligonucleotide probes (consensus and mutant) affixed to Sepharose beads were incubated with nuclear extract and loaded onto a column, and bound proteins were eluted using 500 mM KCl. All flow-through samples and eluates were tested via EMSA (Fig. 5). In the flow-through sample from the column with the mutant oligonucleotide, significant binding was seen in the EMSA; in contrast, in the flow-through from the column with the WT oligonucleotide, no binding was noted, indicating that the protein(s) of interest bound to the oligonucleotide probe containing the HRM. No binding was observed in the 200 mM KCl wash samples from either column. In the 500 mM KCl samples, HRM-binding protein was in the eluate from the WT oligonucleotide column, but no protein binding to HRM was in the eluate from the column with the mutant oligonucleotide. This demonstrates that the protein(s) of interest are enriched in the sample from the WT oligonucleotide column. The 500 mM KCl eluates were therefore examined by mass spectrometry to iden-

**TABLE 1**

**Identity of H₂O2-responsive promoters that were selected for EMSA and functional characterization**

Promoters were selected based on similarity of the promoter HRM to the consensus HRM and proximity of the motif to the start codon (ATG). Gene ID, gene name, and promoter-specific HRMs are shown. Nucleotides that differ from the consensus HRM are underlined and in boldface type. The microarray data originate from Vicente *et al*, (2). Distance upstream of ATG was determined based on the number of nucleotides from position 15 of the motif sequence to the A of ATG.

| Gene ID | Gene name | Sequence | Normalized microarray data | | | Distance upstream of ATG |
|---------|-----------|----------|-----------|-----------|--------|--------------------------|
| | | | Base line | With $H_2O_2$ | Change | |
| | | | | | *-fold* | *nucleotides* |
| Consensus | | AAACCTCAATGAAGA | | | | |
| EHI_029340 | Short chain dehydrogenase family protein | AAACCTCAATG**TT**GA | 0.06 | 0.18 | 3.15 | 58 |
| EHI_134960 | Skp1-related protein | AAACCTCAAT**AAACT** | 34.10 | 68.90 | 2.01 | 44 |
| EHI_176810 | Hypothetical protein | AAACCTCA**TT**GT**TTC**A | 0.10 | 0.87 | 8.68 | 74 |
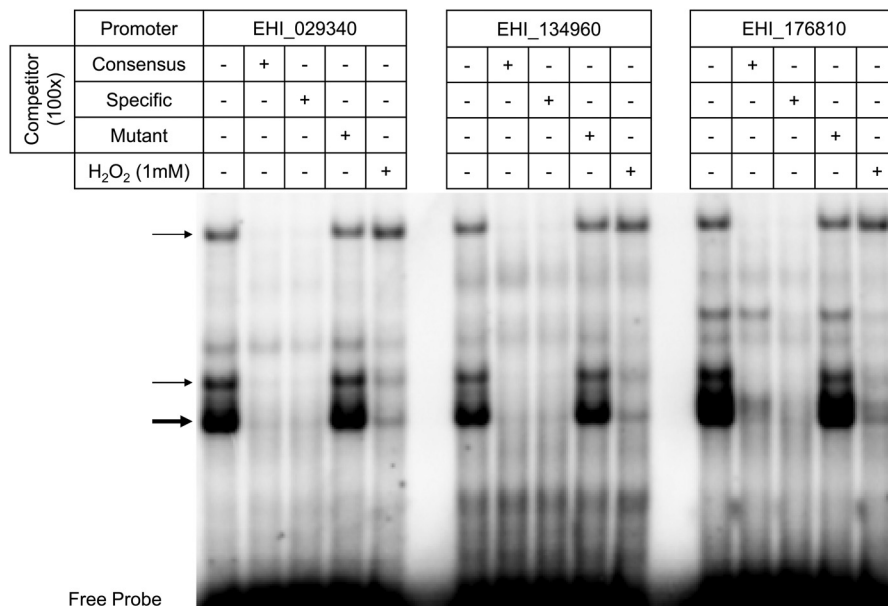


FIGURE 3. **Binding to promoter specific motifs is specific and decreases with H₂O₂.** EMSAs were performed using 5 $\mu$g of crude nuclear extract from untreated or H₂O₂-stressed (1 mM, 1 h) amoebae and $^{32}$P-labeled probes specific to the promoter motif (see Table 1 for sequence). Competition assays were performed using a 100-fold excess of cold competitor of either the consensus probe, probes that perfectly match the relevant promoter motif (specific competitor), or mutant competitor probes in which the nucleotides at positions 7 and 9 are switched (mutant competitor). The *arrows* indicate major bands that exhibit specific binding.

tify which proteins were unique (or enriched) in the WT samples. MS was performed on three independently generated samples from DNA affinity chromatography columns containing WT or mutant oligonucleotides. Overall, MS analysis of the WT and mutant samples resulted in the identification of 549 proteins that met the cut-off criteria (minimum protein, 95%; minimum number of peptides, 1; minimum peptide, 95%) in at least one of the six samples analyzed (supplemental Table 4).

*Direct Isolation of the Oligonucleotide Probe and Protein(s) of Interest*—As an alternative strategy to the DNA affinity columns, we also directly isolated the protein(s)-probe complex from a polyacrylamide gel. To achieve this, we performed three EMSA binding reactions using nuclear extract and an oligonucleotide probe containing the HRM and resolved them on a polyacrylamide gel. The two flanking samples contained radiolabeled probe, and the middle sample contained unlabeled probe. The gel was exposed to film overnight and then aligned on top of the film, allowing the location of the three samples to be determined. A small section of gel was excised that corresponded to the location of the cold probe-protein complex and was analyzed by MS. MS analysis of this sample identified 127 proteins present in the gel slice. Of these, 32 overlapped with

proteins identified in the DNA affinity purification MS analysis (supplemental Table 4).

*Identification of Protein EHI_108720*—Proteins of interest were required to be present in the MS data from all three of the WT oligonucleotide column samples and absent or greatly reduced in the mutant oligonucleotide column samples. Additionally, proteins of interest were required to be present in the MS data from the gel slice. Only one protein fulfilled all criteria. EHI_108720 is a hypothetical protein that is predicted to be a 50-kDa protein consisting of 444 amino acids. With the exception of homologous proteins in *Entamoeba dispar* and *Entamoeba invadens*, BLAST analysis returned no proteins with significant sequence identity. A Pfam search (45) for potential regulatory domains identified a weak match to a helix-turn-helix domain (PF01381) toward the N terminus (*e* value of 0.18), which is of potential interest as a DNA binding domain.

Previously published microarray data show that EHI_108720 is moderately expressed in the pathogenic strain of *E. histolytica*, HM1:IMSS (2), and that mRNA expression levels are not significantly altered in response to H₂O₂, dipropylenetriamine-NONOate, or heat shock or during colonic tissue invasion (2, 29, 45). Additionally, mRNA levels are similar in the avirulent
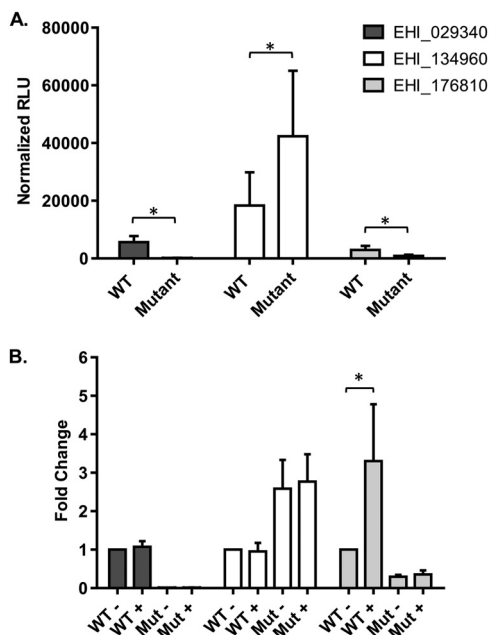
FIGURE 4. **Mutation of the HRM impacts basal luciferase expression levels and H$_2$O$_2$ responsiveness.** Promoters for genes EHI_029340, EHI_134960, and EHI_176810 were cloned into an *E. histolytica* luciferase reporter plasmid. For each construct, site-directed mutagenesis was performed, resulting in the nucleotides at positions 7 and 9 of the HRM being switched. Wild type and mutant promoters were assessed for their ability to drive luciferase expression in transient transfection assays (*A*) or the response of each promoter to H$_2$O$_2$ stress (1 mM, 1 h) (*B*). −, no stress; +, H$_2$O$_2$ stress. -Fold change data from at least three experiments are presented. *, $p < 0.01$. *Error bars*, S.D.
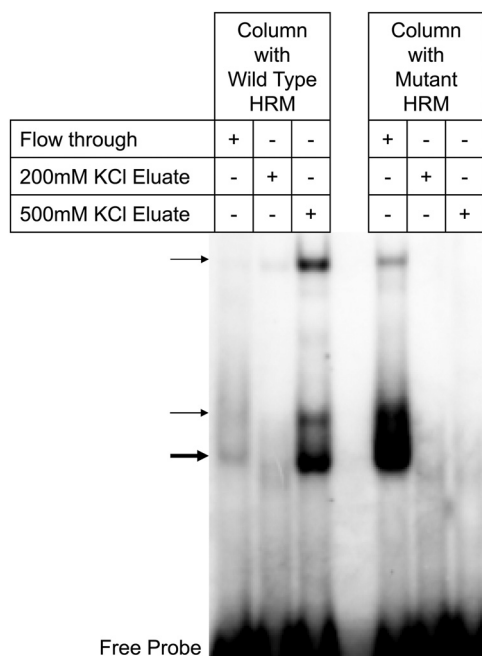


FIGURE 5. **Enrichment of DNA binding protein(s) from *E. histolytica* crude nuclear extracts.** A DNA affinity chromatography assay was performed to enrich DNA-binding protein(s) that bind to the consensus HRM. Nuclear protein was applied to columns with either the wild-type or mutant consensus HRM, washed with 200 mM KCl EMSA binding buffer, and eluted with 500 mM KCl EMSA binding buffer. 5 μl of the samples from the flow-through, 200 mM wash, and 500 mM elution samples were assessed in EMSA using the ^{32}P-labeled consensus probe. The *arrows* indicate major bands that exhibit specific binding.

*E. histolytica* Rahman strain (2). These data indicate that EHI_108720 mRNA remains stably expressed under a wide range of conditions.
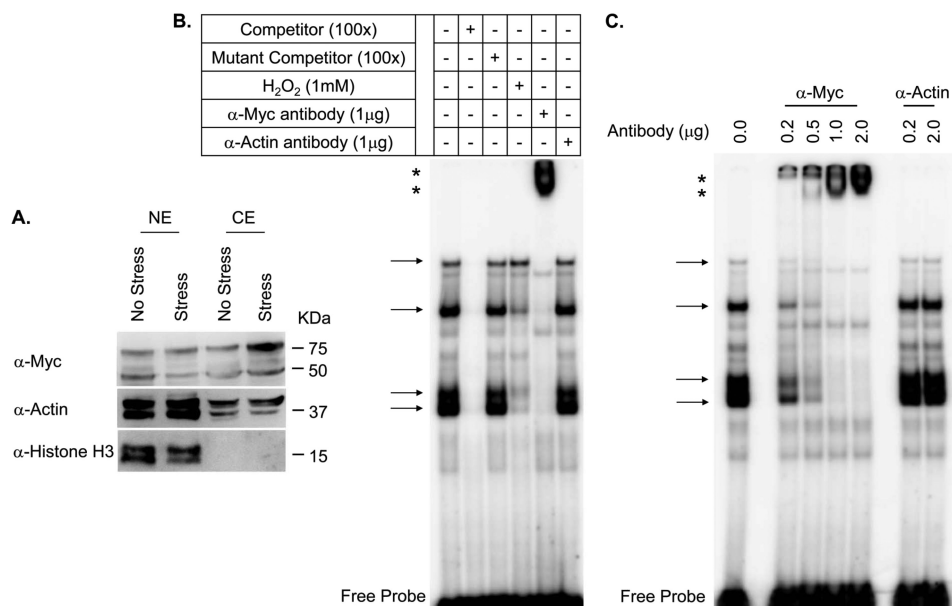
### Characterization of Myc-tagged EHI_108720

In order to characterize EHI_108720, we expressed an N-terminal Myc-tagged version of the protein in *E. histolytica* trophozoites. Confirmation that the gene was overexpressed in this cell line was achieved by performing RT-PCR (Fig. 7*A*). We then isolated cytoplasmic and nuclear enriched fractions from the transfected cell line and probed for the Myc-tagged protein by Western blot analysis (Fig. 6*A*). The EHI_108720 protein was detected at two different sizes, at both the expected 50-kDa size and at 75 kDa, and was present in the nucleus, an important requirement for a potential transcription factor. Additionally, a substantial amount of protein was present in the cytoplasmic fraction. The localization and distribution of EHI_108720 was confirmed using an immunofluorescence assay (data not shown). Cytoplasmic localization is not uncommon for transcription factors; for example, URE3-BP is present in both the nucleus and cytoplasm (46). Changes in the cellular localization of URE3-BP have been linked to a regulatory mechanism for this transcription factor (46). The transfected cell line overexpressing EHI_108720 was exposed to H$_2$O$_2$ stress, and a Western blot was performed to determine if there was any change in abundance, size, or localization of EHI_108720. However, there was no reproducible difference in the size, abundance, or localization of EHI_108720 between samples exposed to H$_2$O$_2$ stress and those from the unstressed parasites (Fig. 6*A*).

### Specific Interaction between EHI_108720 Protein and the HRM

*Overexpressed Myc-tagged EHI_108720 Protein Specifically Binds the HRM*—In order to confirm that the overexpressed protein was able to bind to the HRM motif, nuclear enriched fractions were isolated from EHI_108720 overexpression cell lines (under untreated and stress conditions) and assessed for the presence of HRM-binding protein using EMSA (Fig. 6*B*). Four distinct bands were present, all of which were absent when cold competitor was provided in excess. The addition of mutant competitor did not affect binding, indicating that each of these bands represents a specific interaction between protein and probe. Nuclear extract from amoebae overexpressing HRM-BP and H$_2$O$_2$ stressed showed a decrease in three of the four EMSA bands. This decrease in binding was also observed when using nuclear enriched fractions from wild-type H$_2$O$_2$-stressed amoebae (Fig. 2). Of note, the decreased binding occurs despite the lack of observable differences in protein distribution or quantity observed via Western blot (Fig. 6*A*).

In order to confirm that bands on the EMSA represented a specific interaction between the Myc-tagged EHI_108720 protein and the labeled probe, we performed supershift assays using α-Myc antibody; an α-actin antibody was also used to confirm the specificity of the changes (Fig. 6, *B* and *C*). Increasing concentration of the α-Myc antibody resulted in the four bands decreasing to below detectable levels and resulted in two new bands in the uppermost portion of the gel, representing the supershifted complex. The control α-actin antibody had no effect on binding and did not result in a supershift. Overall, this

FIGURE 6. **Myc-tagged EHI_108720 localizes to both the nucleus and cytoplasm and specifically binds the consensus HRM.** *A*, distribution of Myc-tagged EHI_108720 was assessed via Western blot analysis. $\alpha$-Histone H3 and $\alpha$-actin antibodies were utilized as loading controls between nuclear (*NE*) and cytoplasmic (*CE*) fractions, respectively (stress = 1 mM $H_2O_2$, 1 h). *B*, EMSA analysis demonstrated that overexpressed protein binds to the HRM consensus motif. Competition assays, using a 100-fold excess of cold competitor, and supershift assays ($\alpha$-Myc and $\alpha$-actin antibodies) confirmed the specificity of this interaction. The *arrows* indicate major bands that exhibit specific binding. *, supershifted bands. *C*, all four major bands decrease proportionally with increasing amounts of $\alpha$-Myc antibody, confirming that the Myc-tagged protein is present in each protein-probe complex. $\alpha$-Actin antibody at the lowest and highest concentrations did not impact binding of the four major bands. *, supershifted bands.

analysis confirmed a specific interaction between the Myc-tagged EHI_108720 and the consensus HRM motif.

*Knockdown of EHI_108720 Leads to Abrogation of Binding to the HRM*—Knockdown of EHI_108720 was achieved using an RNAi-based method with significant reduction in gene-specific mRNA compared with controls (Fig. 7*A*). EMSA analysis with nuclear extract from the knockdown cell line demonstrated decreased binding to the labeled probe despite equal binding to a control probe (Fig. 7, *B* and *C*). Taken together, the genetic manipulation of EHI_108720, by overexpression and knockdown experiments, and associated changes in binding as shown by EMSA confirm that EHI_108720 is an HRM-BP.

### Genetic Manipulation of EHI_108720 Influences Gene Expression from an $H_2O_2$-responsive Promoter

Because a specific interaction between the HRM and EHI_108720 was confirmed, we addressed the impact of manipulating HRM-BP on expression from the $H_2O_2$-responsive promoter of EHI_176810. RT-PCR analysis confirmed that in the untransfected control cell line (*UT*) the levels of EHI_176810 mRNA increased following the application of $H_2O_2$ stress (Fig. 8*A*). In the HRM-BP overexpression cell line, EHI_176810 mRNA levels were greater than those in the control sample, but $H_2O_2$ exposure had no effect on mRNA levels. In the HRM-BP knockdown cell line, the mRNA levels of EHI_176810 still demonstrated an increase following stress.

Having ascertained the effects of manipulating HRM-BP on the expression of EHI_176810 mRNA, we next sought to determine the effects of HRM-BP on reporter protein expression. To achieve this, the EHI_176810 promoter-luciferase fusion constructs (WT and mutant) were transiently transfected into either HRM-BP overexpression, HRM-BP knockdown, or
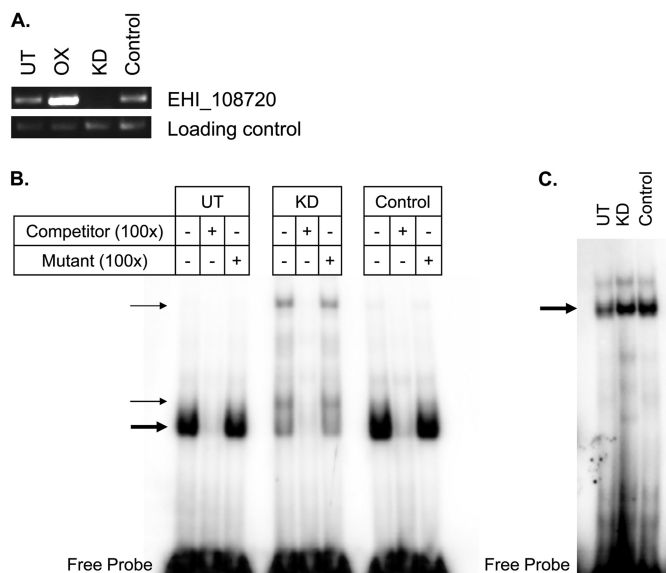


FIGURE 7. **Knockdown of EHI_108720 results in decreased binding to the consensus HRM.** *A*, overexpression and knockdown of the putative HRM-BP, EHI_108720, was assessed using RT-PCR and compared with untransfected and control cell lines. *B*, nuclear enriched fractions from untransfected, knockdown, and control cell lines were assessed for specific binding to the consensus HRM. Competition assays were performed using a 100-fold excess of cold competitor of either the consensus probe or mutant probe. *C*, an alternate oligonucleotide probe was employed as a loading control to ensure that an equal amount of protein was loaded in each sample. The *arrows* indicate major bands that exhibit specific binding. *UT*, untransfected control; *OX*, over-expression cell line; *KD*, knockdown cell line.

untransfected control cell lines. Luciferase expression levels were determined for all cell lines before and after exposure to $H_2O_2$ (Fig. 8*B*). As expected, oxidative stress resulted in an ~3-fold increase in luciferase levels from the WT promoter in the control cells ($p < 0.05$), similar to previous observations
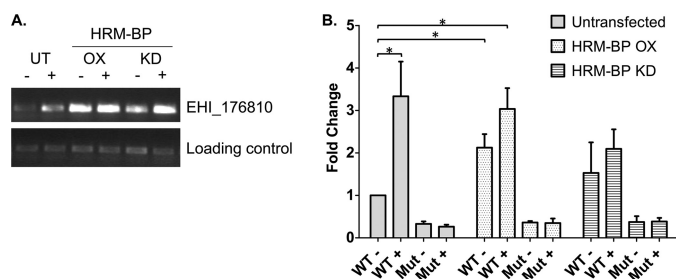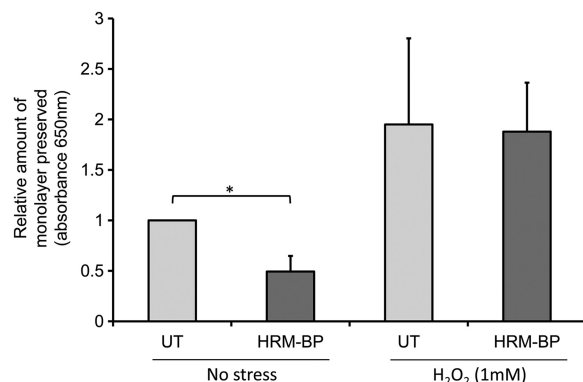
FIGURE 8. **Overexpression or knockdown of EHI_108720 influences gene expression from an $H_2O_2$-responsive promoter.** *A*, the effects of HRM-BP manipulation on EHI_176810 mRNA levels were assessed via RT-PCR. Untransfected control, overexpression, and knockdown cell lines were examined before and after exposure to $H_2O_2$ (1 mM, 1 h). *B*, untransfected, overexpression, and knockdown cell lines were transiently transfected with the EHI_176810 promoter-luciferase construct. Untreated or $H_2O_2$-exposed (1 mM, 1 h) parasites were assessed for luciferase activity. Fold change data from five experiments are presented. *, *p* value < 0.05. *Error bars*, S.E.; *UT*, untransfected control; *KD*, knockdown; *OX*, overexpression. −, no stress; +, $H_2O_2$ stress.



FIGURE 9. **Overexpression of EHI_108720 increases parasite virulence.** The cytotoxicity of HRM-BP overexpression and control cell lines (*UT*) was determined using a monolayer destruction assay. $H_2O_2$-stressed (1 mM, 1 h) or untreated parasites were added to confluent CHO monolayers. The level of monolayer destruction was assessed using 0.1% methylene blue, dye extraction, and absorbance at 650 nm. Relative absorbance levels are shown for four experiments performed in duplicate. *, *p* < 0.05. *Error bars*, S.D.

(Fig. 4*B*). In the HRM-BP-overexpressing cell line, there was an ~2-fold increase in luciferase levels from the unstressed WT promoter compared with the luciferase levels observed in the control cell line ($p < 0.05$). As a consequence of increased basal luciferase expression, the effect of $H_2O_2$ exposure was muted and no longer significantly different (~1.4-fold increase) (Fig. 8*B*). In the HRM-BP knockdown cell line, the basal levels of luciferase expression were not significantly different from the basal levels seen in the control cell line. Additionally, exposure to $H_2O_2$ no longer significantly increased levels of the reporter gene (Fig. 8*B*). In all three cell lines, transfection with the mutant promoter constructs resulted in decreased luciferase expression compared with the WT counterparts and were unaffected by exposure to $H_2O_2$.

These data demonstrate that manipulation of the HRM-BP similarly alters both mRNA and reporter gene protein levels. Importantly, overexpression of this protein significantly raises basal transcription levels and negates the effect of exposure to hydrogen peroxide. Given that overexpression increased basal luciferase expression, it could be hypothesized that knockdown of EHI_108720 would negatively affect basal levels. However, this was not observed, which may be due to incomplete knockdown of EHI_108720 and low levels of protein persisting, as suggested by the EMSA analysis using extract from the knockdown cell line (Fig. 7*B*). However, knockdown of EHI_108720 did negate the ability of the WT promoter to significantly increase luciferase expression in response to $H_2O_2$ exposure (Fig. 8*B*). Results from the mutant promoters confirm that EHI_108720 requires the presence of an intact HRM to influence reporter gene expression from this promoter. Taken together, these findings confirm that EHI_108720 impacts gene expression and that it mediates its effect via the HRM.

### Overexpression of HRM-BP Increases Parasite Virulence

Given that overexpression of the HRM-BP has the potential to impact basal gene expression via the HRM, we examined this cell line for phenotypic changes. Comparisons against a control cell line (maintained at an identical drug concentration) determined that overexpression of the HRM-BP had no effect on

growth rate or on cell survival following exposure to $H_2O_2$ (data not shown).

Monolayer destruction assays were performed to determine if overexpression of the HRM-BP resulted in changes in cytotoxicity (Fig. 9). Compared with the control cell line, the HRM-BP overexpression cell line destroyed a significantly greater amount of the target cells ($p \leq 0.05$), indicating an increase in virulence. Exposure to $H_2O_2$ resulted in decreased cytotoxicity in both the control and the HRM-BP overexpression cell lines when compared with the unstressed parasites.

Overall, these data suggest not only that overexpression of the HRM-BP affects basal expression levels of a subset of genes but that these changes result in increased cytotoxicity and that HRM-BP significantly impacts the virulence potential of *E. histolytica*.

## DISCUSSION

Invasive pathogens are frequently exposed to reactive oxygen species as a mechanism of host defense, and an organism's ability to survive this aspect of host immunity is critical to establishing infection. In many systems, the transcriptional machinery that regulates gene expression in response to ROS is well characterized (12, 13, 47–52). In the pathogenic protist *E. histolytica*, increased virulence has been linked to the parasite's ability to survive exposure to ROS (21), and the transcriptional response to hydrogen peroxide stress has been characterized (2). However, the regulatory pathways involved in controlling gene expression in response to ROS have not been identified, and homologues of transcription factors that respond to ROS are absent from the amoebic genome (30, 31). In this study, we utilized a bioinformatics approach to identify promoter motifs enriched in genes up-regulated in response to hydrogen peroxide. We identified AAACCTCAATGAAGA as a hydrogen peroxide regulatory motif, used EMSA to show that amoebic nuclear proteins bind specifically to this motif, demonstrated that substitution of two core nucleotides (C7 and A9) is sufficient to abrogate DNA-protein binding, and established the biological importance of the HRM using reporter gene assays. Furthermore, we employed DNA affinity chromatography and

mass spectrometry to identify the HRM-binding protein (EHI_108720). Confirmation of this interaction and the biological significance of HRM-BP was achieved by assessing amoebic cell lines in which the protein was either overexpressed or knocked down. Overall, this work represents the first identification of a transcription factor that controls the coordinated regulation of oxidative stress response genes in *E. histolytica* and has important ramifications for understanding the molecular basis of stress response in an important human pathogen.

The mechanism by which the HRM-binding protein (EHI_108720) regulates gene expression through the HRM is not yet known, but our data suggest several intriguing features. Functional analysis of three HRM-containing promoters determined that this motif asserts activating or repressing effects, with divergent roles noted in a promoter-specific manner. Of interest, a similar observation was noted with mutational analysis of the *E. histolytica* URE3 motif in the promoters of the *hgl5* and ferredoxin 1 genes (37, 44). The variable role that the HRM plays in modulating transcription could be explained by a number of factors, including the attributes of the surrounding DNA, small changes in the nucleotide composition of the HRM itself, or properties of the HRM-binding protein. From the original microarray study, we know that the basal transcription levels and the magnitude of the response to stress varies for each promoter. This indicates that other components, specific to each promoter, influence gene expression. It is also plausible that the small changes seen in the nucleotide composition of the promoter-specific HRMs influence the role it plays in regulating transcription. This scenario is exemplified by the dual role observed for the transcription factor Pit-1, where Pit-1 activity is influenced by the presence or absence of an additional two nucleotides within its DNA binding motif (53). The conformational change induced in the Pit-1 dimer by the inclusion of the two nucleotides enables the recruitment of the transcriptional co-repressor N-CoR and the reversal of its activating role (53).

Alternatively, the variable outcome of mutating the HRM could be dependent on the intrinsic properties of the protein and/or the regulatory co-factors that it recruits. Given that ~65% of the genes annotated in the *E. histolytica* genome encode hypothetical proteins (31), it was not surprising that a protein of unknown function was identified as the HRM-BP. Unfortunately, the lack of identifiable domains makes it difficult to immediately predict how this protein differentially regulates transcription. Overexpression of the HRM-BP revealed that binding to the consensus motif results in four distinct bands, each of which contains the HRM-BP. Whether these bands represent concatemers of the HRM-BP or whether the increased size is due to additional protein co-factors being recruited to bind to the HRM is not known at present. Therefore, whether HRM-BP activates or represses expression from a given promoter may depend upon the components of the protein complex bound at each promoter motif.

Given that expression or localization of HRM-BP does not change in response to oxidative stress, it is not readily apparent how this protein is regulated in response to hydrogen peroxide. In *E. coli* and yeast, the OxyR and YAP-1 proteins direct increased expression of a subset of genes in response to hydrogen peroxide (1, 6, 7, 12, 13). In *Bacillus subtilis*, one of the key transcription factors involved in up-regulating gene expression in response to hydrogen peroxide is PerR, a transcriptional repressor (5, 48). In all three cases, the protein itself is altered in response to hydrogen peroxide exposure (increased DNA binding affinity (OxyR) (14, 15), increased nuclear retention (YAP-1) (16), and impaired DNA binding (PerR) (54)). In each situation, the protein alteration was key to the transcriptional regulation that occurs in these organisms. Given the lack of change in mRNA levels observed for EHI_108720, the multiple bands observed in EMSA analysis, and the decreased binding seen following exposure to stress, it is interesting to speculate that changes in the HRM-BP protein itself may be central to its behavior.

In addition to understanding how HRM-BP is regulated, it is important to consider the role it plays in pathogenesis. The initial microarray study looking at global transcriptional changes in response to $H_2O_2$ observed that many genes known to be involved in mitigating the effects of ROS exposure were already highly expressed and were not further affected by stress. Consistent with this observation, the phenotypic studies performed in this study demonstrated that overexpression of HRM-BP does not confer increased resistance to $H_2O_2$. However, the overexpression of HRM-BP does increase the cytotoxicity of *E. histolytica* and suggests that at least a portion of the genes regulated in response to $H_2O_2$ exposure are involved in other aspects of amoebic biology, such as virulence.

HRM-BP represents a novel transcription factor that is involved in a fundamentally important aspect of *E. histolytica* biology, and its identification is a significant first step in understanding how the parasite coordinates gene expression in response to hydrogen peroxide stress. The link between susceptibility to ROS and virulence underscores the importance of this area of research. Further characterization of this protein and its binding partners will help to fully elucidate its role in parasite virulence.

## REFERENCES

1. Chen, D., Toone, W. M., Mata, J., Lyne, R., Burns, G., Kivinen, K., Brazma, A., Jones, N., and Bähler, J. (2003) Global transcriptional responses of fission yeast to environmental stress. *Mol. Biol. Cell* **14,** 214–229
2. Vicente, J. B., Ehrenkaufer, G. M., Saraiva, L. M., Teixeira, M., and Singh, U. (2009) *Entamoeba histolytica* modulates a complex repertoire of novel genes in response to oxidative and nitrosative stresses. Implications for amebic pathogenesis. *Cell Microbiol.* **11,** 51–69
3. Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D., and Brown, P. O. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11,** 4241–4257
4. Li, H., Singh, A. K., McIntyre, L. M., and Sherman, L. A. (2004) Differential gene expression in response to hydrogen peroxide and the putative PerR regulon of *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* **186,** 3331–3345
5. Mostertz, J., Scharf, C., Hecker, M., and Homuth, G. (2004) Transcriptome and proteome analysis of *Bacillus subtilis* gene expression in response to superoxide and peroxide stress. *Microbiology* **150,** 497–512
6. Zheng, M., Wang, X., Templeton, L. J., Smulski, D. R., LaRossa, R. A., and Storz, G. (2001) DNA microarray-mediated transcriptional profiling of

the *Escherichia coli* response to hydrogen peroxide. *J. Bacteriol.* **183,** 4562–4570

7. Godon, C., Lagniel, G., Lee, J., Buhler, J. M., Kieffer, S., Perrot, M., Boucherie, H., Toledano, M. B., and Labarre, J. (1998) The H$_2$O$_2$ stimulon in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **273,** 22480–22489

8. Halliwell, B., Gutteridge, J. M. C. (2007) *Free Radicals in Biology and Medicine*, Oxford University Press, Oxford

9. Justino, M. C., Vicente, J. B., Teixeira, M., and Saraiva, L. M. (2005) New genes implicated in the protection of anaerobically grown *Escherichia coli* against nitric oxide. *J. Biol. Chem.* **280,** 2636–2643

10. Vicente, J. B., Tran, V., Pinto, L., Teixeira, M., and Singh, U. (2012) A detoxifying oxygen reductase in the anaerobic protozoan *Entamoeba histolytica*. *Eukaryot. Cell* **11,** 1112–1118

11. Vandenbroucke, K., Robbens, S., Vandepoele, K., Inzé, D., Van de Peer, Y., and Van Breusegem, F. (2008) Hydrogen peroxide-induced gene expression across kingdoms. A comparative analysis. *Mol. Biol. Evol.* **25,** 507–516

12. Stephen, D. W., Rivers, S. L., and Jamieson, D. J. (1995) The role of the YAP1 and YAP2 genes in the regulation of the adaptive oxidative stress responses of *Saccharomyces cerevisiae*. *Mol. Microbiol.* **16,** 415–423

13. Christman, M. F., Storz, G., and Ames, B. N. (1989) OxyR, a positive regulator of hydrogen peroxide-inducible genes in *Escherichia coli* and *Salmonella typhimurium*, is homologous to a family of bacterial regulatory proteins. *Proc. Natl. Acad. Sci. U.S.A.* **86,** 3484–3488

14. Zheng, M., Aslund, F., and Storz, G. (1998) Activation of the OxyR transcription factor by reversible disulfide bond formation. *Science* **279,** 1718–1721

15. Choi, H., Kim, S., Mukhopadhyay, P., Cho, S., Woo, J., Storz, G., and Ryu, S. E. (2001) Structural basis of the redox switch in the OxyR transcription factor. *Cell* **105,** 103–113

16. Kuge, S., Jones, N., and Nomoto, A. (1997) Regulation of yAP-1 nuclear localization in response to oxidative stress. *EMBO J.* **16,** 1710–1720

17. Georgiou, G. (2002) How to flip the (redox) switch. *Cell* **111,** 607–610

18. Stanley, S. L., Jr. (2003) Amoebiasis. *Lancet* **361,** 1025–1034

19. Mattern, C. F., Keister, D. B., and Caspar, P. A. (1978) Experimental amebiasis. III. A rapid *in vitro* assay for virulence of *Entamoeba histolytica*. *Am. J. Trop Med. Hyg.* **27,** 882–887

20. Davis, P. H., Schulze, J., and Stanley, S. L., Jr. (2007) Transcriptomic comparison of two *Entamoeba histolytica* strains with defined virulence phenotypes identifies new virulence factor candidates and key differences in the expression patterns of cysteine proteases, lectin light chains, and calmodulin. *Mol. Biochem. Parasitol.* **151,** 118–128

21. Davis, P. H., Zhang, X., Guo, J., Townsend, R. R., and Stanley, S. L., Jr. (2006) Comparative proteomic analysis of two *Entamoeba histolytica* strains with different virulence phenotypes identifies peroxiredoxin as an important component of amoebic virulence. *Mol. Microbiol.* **61,** 1523–1532

22. MacFarlane, R. C., and Singh, U. (2006) Identification of differentially expressed genes in virulent and nonvirulent *Entamoeba* species. Potential implications for amebic pathogenesis. *Infect. Immun.* **74,** 340–351

23. Bruchhaus, I., Richter, S., and Tannich, E. (1997) Removal of hydrogen peroxide by the 29-kDa protein of *Entamoeba histolytica*. *Biochem. J.* **326,** 785–789

24. Pearson, R. J., and Singh, U. (2010) Approaches to characterizing *Entamoeba histolytica* transcriptional regulation. *Cell Microbiol.* **12,** 1681–1690

25. Meneses, E., Cárdenas, H., Zárate, S., Brieba, L. G., Orozco, E., López-Camarillo, C., and Azuara-Liceaga, E. (2010) The R2R3 Myb protein family in *Entamoeba histolytica*. *Gene* **455,** 32–42

26. de Dios-Bravo, G., Luna-Arias, J. P., Riverón, A. M., Olivares-Trejo, J. J., López-Camarillo, C., and Orozco, E. (2005) Entamoeba histolytica TATA-box binding protein binds to different TATA variants in vitro. *FEBS J.* **272,** 1354–1366

27. Gilchrist, C. A., Holm, C. F., Hughes, M. A., Schaenman, J. M., Mann, B. J., and Petri, W. A., Jr. (2001) Identification and characterization of an *Entamoeba histolytica* upstream regulatory element 3 sequence-specific DNA-binding protein containing EF-hand motifs. *J. Biol. Chem.* **276,** 11838–11843

28. Gilchrist, C. A., Leo, M., Line, C. G., Mann, B. J., and Petri, W. A., Jr. (2003)

Calcium modulates promoter occupancy by the *Entamoeba histolytica* Ca$^{2+}$-binding transcription factor URE3-BP. *J. Biol. Chem.* **278,** 4646–4653

29. Hackney, J. A., Ehrenkaufer, G. M., and Singh, U. (2007) Identification of putative transcriptional regulatory networks in *Entamoeba histolytica* using Bayesian inference. *Nucleic Acids Res.* **35,** 2141–2152

30. Loftus, B., Anderson, I., Davies, R., Alsmark, U. C., Samuelson, J., Amedeo, P., Roncaglia, P., Berriman, M., Hirt, R. P., Mann, B. J., Nozaki, T., Suh, B., Pop, M., Duchene, M., Ackers, J., Tannich, E., Leippe, M., Hofer, M., Bruchhaus, I., Willhoeft, U., Bhattacharya, A., Chillingworth, T., Churcher, C., Hance, Z., Harris, B., Harris, D., Jagels, K., Moule, S., Mungall, K., Ormond, D., Squares, R., Whitehead, S., Quail, M. A., Rabbinowitsch, E., Norbertczak, H., Price, C., Wang, Z., Guillén, N., Gilchrist, C., Stroup, S. E., Bhattacharya, S., Lohia, A., Foster, P. G., Sicheritz-Ponten, T., Weber, C., Singh, U., Mukherjee, C., El-Sayed, N. M., Petri, W. A., Jr., Clark, C. G., Embley, T. M., Barrell, B., Fraser, C. M., and Hall, N. (2005) The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433,** 865–868

31. Lorenzi, H. A., Puiu, D., Miller, J. R., Brinkac, L. M., Amedeo, P., Hall, N., and Caler, E. V. (2010) New assembly, reannotation and analysis of the *Entamoeba histolytica* genome reveal new genomic features and protein content information. *PLoS Negl. Trop Dis.* **4,** e716

32. Ehrenkaufer, G. M., Hackney, J. A., and Singh, U. (2009) A developmentally regulated Myb domain protein regulates expression of a subset of stage-specific genes in *Entamoeba histolytica*. *Cell Microbiol.* **11,** 898–910

33. Bailey, T. L., and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2,** 28–36

34. Bailey, T. L., and Gribskov, M. (1998) Combining evidence using *p*-values. Application to sequence homology searches. *Bioinformatics* **14,** 48–54

35. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo. A sequence logo generator. *Genome Res.* **14,** 1188–1190

36. Diamond, L. S., Harlow, D. R., and Cunnick, C. C. (1978) A new medium for the axenic cultivation of *Entamoeba histolytica* and other *Entamoeba*. *Trans. R. Soc. Trop. Med. Hyg.* **72,** 431–432

37. Gilchrist, C. A., Mann, B. J., and Petri, W. A., Jr. (1998) Control of ferredoxin and Gal/GalNAc lectin gene expression in *Entamoeba histolytica* by a cis-acting DNA sequence. *Infect. Immun.* **66,** 2383–2386

38. Purdy, J. E., Mann, B. J., Pho, L. T., and Petri, W. A., Jr. (1994) Transient transfection of the enteric parasite *Entamoeba histolytica* and expression of firefly luciferase. *Proc. Natl. Acad. Sci. U.S.A.* **91,** 7099–7103

39. Smith, A. J., Chudnovsky, L., Simoes-Barbosa, A., Delgadillo-Correa, M. G., Jonsson, Z. O., Wohlschlegel, J. A., and Johnson, P. J. (2011) Novel core promoter elements and a cognate transcription factor in the divergent unicellular eukaryote *Trichomonas vaginalis*. *Mol. Cell Biol.* **31,** 1444–1458

40. Meng, Z., Camalier, C. E., Lucas, D. A., Veenstra, T. D., Beck, G. R., Jr., and Conrads, T. P. (2006) Probing early growth response 1 interacting proteins at the active promoter in osteoblast cells using oligoprecipitation and mass spectrometry. *J. Proteome Res.* **5,** 1931–1939

41. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1,** 2856–2860

42. Ehrenkaufer, G. M., Haque, R., Hackney, J. A., Eichinger, D. J., and Singh, U. (2007) Identification of developmentally regulated genes in *Entamoeba histolytica*. Insights into mechanisms of stage conversion in a protozoan parasite. *Cell Microbiol.* **9,** 1426–1444

43. Rastew, E., Vicente, J. B., and Singh, U. (2012) Oxidative stress resistance genes contribute to the pathogenic potential of the anaerobic protozoan parasite, *Entamoeba histolytica*. *Int. J. Parasitol.* **42,** 1007–1015

44. Purdy, J. E., Pho, L. T., Mann, B. J., and Petri, W. A., Jr. (1996) Upstream regulatory elements controlling expression of the *Entamoeba histolytica* lectin. *Mol. Biochem. Parasitol.* **78,** 91–103

45. Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E. L., Eddy, S. R., Bateman, A., and Finn, R. D. (2012) The Pfam protein families database. *Nucleic Acids Res.* **40,** D290–D301

46. Moreno, H., Linford, A. S., Gilchrist, C. A., and Petri, W. A., Jr. (2010)

Phospholipid-binding protein EhC2A mediates calcium-dependent translocation of transcription factor URE3-BP to the plasma membrane of *Entamoeba histolytica. Eukaryot. Cell* **9,** 695–704

47. Dinamarco, T. M., Almeida, R. S., de Castro, P. A., Brown, N. A., dos Reis, T. F., Ramalho, L. N., Savoldi, M., Goldman, M. H., and Goldman, G. H. (2012) Molecular characterization of the putative transcription factor SebA involved in virulence in *Aspergillus fumigatus. Eukaryot. Cell* **11,** 518–531

48. Bsat, N., Herbig, A., Casillas-Martinez, L., Setlow, P., and Helmann, J. D. (1998) *Bacillus subtilis* contains multiple Fur homologues. Identification of the iron uptake (Fur) and peroxide regulon (PerR) repressors. *Mol. Microbiol.* **29,** 189–198

49. Masloboeva, N., Reutimann, L., Stiefel, P., Follador, R., Leimer, N., Hennecke, H., Mesa, S., and Fischer, H. M. (2012) Reactive oxygen species-inducible ECF $\sigma$ factors of *Bradyrhizobium japonicum. PLoS One* **7,** e43421

50. Ning, J. L., Mo, L. W., and Lai, X. N. (2010) Low- and high-dose hydrogen peroxide regulation of transcription factor NF-E2-related factor 2. *Chin. Med. J.* **123,** 1063–1069

51. Schmidt, K. N., Amstad, P., Cerutti, P., and Baeuerle, P. A. (1995) The roles of hydrogen peroxide and superoxide as messengers in the activation of transcription factor NF-$\kappa$B. *Chem. Biol.* **2,** 13–22

52. Jacquier-Sarlin, M. R., and Polla, B. S. (1996) Dual regulation of heat-shock transcription factor (HSF) activation and DNA-binding activity by $H_2O_2$. Role of thioredoxin. *Biochem. J.* **318,** 187–193

53. Scully, K. M., Jacobson, E. M., Jepsen, K., Lunyak, V., Viadiu, H., Carrière, C., Rose, D. W., Hooshmand, F., Aggarwal, A. K., and Rosenfeld, M. G. (2000) Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science* **290,** 1127–1131

54. Herbig, A. F., and Helmann, J. D. (2001) Roles of metal ions and hydrogen peroxide in modulating the interaction of the *Bacillus subtilis* PerR peroxide regulon repressor with operator DNA. *Mol. Microbiol.* **41,** 849–859